VLAIO-TETRA
Machine Vision for Quality Control
Case 5 - Determination of fat content in donuts after frying

Contents

1 Introduction

In the food processing industry, ensuring high-quality standards is becoming increasingly challenging due to more stringent safety regulations and customer demands, as well as the limitations of manual quality inspections. To address this, computer vision technology has been increasingly used for inline quality control, offering high-speed processing and repeatability. The near-instant result of a computer vision-based prediction creates a large advantage over current wet chemistry techniques, by enabling faster responses on deviating quality metrics but also increasing the total share of reviewed products in comparison to the full production volume.

However, the implementation of computer vision-based algorithms for natural products still lags compared to other industrial sectors. The core cause of this issue lies in the complex visual structure and heterogeneity of textures and colors of these natural products.

The objective of this study is to investigate the potential of machine learning-based regression using features extracted from near-infrared (NIR) hyperspectral imaging for the non-invasive and inline prediction of fat and moisture concentration in bakery products. This is motivated by different factors. First of all, European law mandates accurate labeling of ingredient quantities, and this regulation extends to bakery products. The production process of bakery products typically results in slight fluctuations in raw materials and ingredient concentration during the process. The tolerance for fat is set to 20% for products containing 10-40g of fat per 100g [1]. However, these fluctuations can significantly impact the flavor, aroma, texture, and calorie content of bakery products, resulting in a different taste experience for consumers. Therefore, there is a need for even more stringent quantity requirements than those mandated by European law. Furthermore, the conventional approaches typically consist of manual, destructive tests, due to which only part of the products can be inspected, which could be improved by adopting an inline, non-invasive quality control methodology.

Traditional RGB cameras are not suitable for capturing subtle variations in natural products that the human sensory organs and brain are evolutionary trained on. Therefore, hyperspectral imaging is employed to capture the impact on a large range of spectral bands in visible and near-infrared light. For the case under consideration, a subsample of the full NIR hyperspectral spectrum contains the absorption response of moisture and fat, and specific wavelength bands exhibit a higher response variation based on the molecular structure and chemical concentration of various components. The focus of this work is on investigating the correlation between these wavelength bands and production process parameters, including shortening fat concentration, proofing duration, and frying duration.

2 Literature study

Recent studies have shown that infrared hyperspectral analysis could be a promising technique for non-invasive and inline fat content determination in bakery products. Infrared hyperspectral analysis involves the use of an infrared spectrometer to measure the absorption of light by molecules in the sample. The technique can provide information about the chemical composition of the sample and can be used to determine the fat content.

In past studies, NIR spectroscopy and Fourier-transform NIR spectroscopy have been proposed to measure content characteristics of wheat, wheat dough and bakery products. Mishra et al. (2012) [2] investigated the use of infrared hyperspectral analysis for rapid assessment of quality change and insect infestation in stored wheat grain using Fourier transform-NIR Spectroscopy. The study showed that the technique was accurate and reliable, with a R^2 of 0.901, 0.938 and 0.895 respectively for the prediction of moisture, protein and uric acid. Similarly, a study by Szigedi et al. (2011) [3] used Fourier-transform NIR spectroscopy to determine protein, lipid and sugar contents

of bakery products. Based on the results, the technique proved to be accurate and could be used as a method for assessing the quality of wheat flour with a correlation of 99.07 for lipid determination.

Gomes et al. (2020) [4] employed hyperspectral imaging (HSI) to identify olive varieties from satellite hyperspectral sensors. It succeeded in successfully visualizing the distinction between varieties with an overall accuracy of 81.23. Similar work is presented by Jiang et al. (2020) [5] where the researchers predicted the leaf lard adulteration in minced pork by partial least squares regression on hyperspectral data. This resulted in a correlation of 0.98 on a selected subset of the full spectra.

This case investigates the potential of machine learning-based regression models and corresponding feature selection techniques to improve upon the existing approaches for determining fat and moisture levels in bakery products. In combination with the latest advancements in hyperspectral imaging and image processing on the broad and well-studied domain of spectroscopy on bakery products for ingredient content determination, this results in an alternative technique for accurate, inline and non-invasive measurements in high-volume production processes.

3 Methodology

This section details the proposed methodology for fat and moisture concentration prediction based on NIR hyperspectral images. A first important aspect is the preprocessing of the raw hyperspectral data. This includes the correction of variations in sensor sensitivity, the alignment of the spectral bands and the removal of unwanted variations. Also a number of transformations are applied on the original preprocessed dataset. More specifically, its first and second derivatives are considered to exploit small variations between the wavelengths for the prediction. Finally, this is served as input for a Partial Least Squares regression and a Least- Squares Support Vector Machine to predict the fat and moisture concentration.

3.1 Spectral transformation

In order to account for variations in the sensitivity of the sensor and obtain accurate image measurements, a first step is to apply a flat-field correction method on the raw images [6–8]. This method involves correcting the raw image R using images of the white F and dark D references taken prior to the start of data collection. The process of correcting the raw image involves calculating the corrected image C, which is achieved according to (1). The flat-field correction method ensures that any variations in sensor sensitivity are accounted for, resulting in a more accurate and reliable image. The use of this correction method is critical in image analysis applications, including the analysis of the samples in this study.

$$C = \frac{(R - D)}{(F - D} \tag{1}$$

Two major factors of variability between different spectral images were anticipated by image preprocessing: reflection variability between images and irradiance differences for all measured spectra. In order to obtain accurate and reliable measurements, the raw spectra underwent several signal preprocessing techniques aimed at identifying the most effective method for aligning the spectral bands to equal magnitude and offset. Two widely used preprocessing techniques, Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV), were implemented and employed in this study. MSC is a data preprocessing technique commonly used in spectroscopy to remove the effect of scattering and interference, such as moisture or other impurities from the spectra. A linear regression is fitted for each spectrum against a reference spectrum by least squares, as shown

in (2). This reference spectrum is typically generated by averaging the spectra of a set of standard samples with similar scattering and interference characteristics [9]. The MSC-corrected spectra are determined by (3).

$$X_i = a_i + b_i X_m \tag{2}$$

$$X_i^{msc} = \frac{(X_i - a_i)}{b_i} \tag{3}$$

On the other hand, the SNV technique involves normalizing the data by subtracting the mean and dividing by the standard deviation, with the aim of removing any unwanted variations, as shown in (4). Unlike MSC, SNV does not require a reference spectrum [10].

$$X_i^{snv} = \frac{(X_i - \overline{X_i})}{\sigma_i} \tag{4}$$

By utilizing both MSC and SNV, this study was able to effectively preprocess the raw data and identify the optimal method for extracting the features needed for accurate determination of fat content.

As a data transformation preprocessing step, the first and second derivative of the data are added and compared to the non-derived dataset. These transformations allow to focus on the small variations between the wavelengths.

In order to assess the quality of the data and identify potential outliers, two measures were calculated for each data series: the mean squared error (MSE) to the mean and the Mahalanobis distance, respectively shown in (5) and (6). The MSE provides a measure of the average difference between the data points and the mean value, while the Mahalanobis Distance takes into account the covariance between the variables in the data set.

$$MSE = \frac{1}{n}\sum_{n=i}^{n}(X_i - \overline{X_i})^2 \tag{5}$$

$$MD = \sqrt{(X - \overline{X})V^{-1}(X - \overline{X})^\top} \tag{6}$$

To identify and remove outliers from the data, an empirically-determined filtering threshold was applied to the MSE and Mahalanobis distance metrics. Specifically, any data points with an MSE greater than 0.02 or a Mahalanobis distance greater than 0.5 were flagged as outliers and removed from the data set. By filtering the data in this manner, the study was able to ensure the reliability and accuracy of the results, as well as minimize the impact of any spurious or erroneous data points.

3.2 Machine Learning-based regression

The task of predicting the moisture and fat content of the samples in the dataset relies on an approach in which several supervised learning methods were tested in order to obtain an accurate regression model. Given the strengths and weaknesses of each algorithm and how well they fit both the dataset and problem, a set of two algorithms was retained and compared, namely Partial Least Squares regression (PLS-R) and Least-Squares Support Vector Machine (LS-SVM). PLS regression is the most commonly used method in near-infrared spectroscopy, as it is an efficient linear multivariate algorithm for spectral data. However, PLS may be limited for regression on more complex data such as noisy signals, low concentrations, and non-linear correlations [11–13].

The algorithm involves the vector Y, which represents the properties as determined by reference methods, the matrix X, which is formed by the calibration spectra vectors, and the vector b, which represents the predicted values of the unknown quantities, as shown in Equation 7.

$$\overrightarrow{Y} = X \cdot \overrightarrow{b} \tag{7}$$

The Least Squares Support Vector Machine (LS-SVM) is a kernel-based learning method that, as the name indicates, is a least-squares version of SVMs. It solves a set of linear equations instead of a convex quadratic programming problem for classical SVMs [14].

3.3 Evaluation

Two measures were used for scoring the overall performance of the model. Root mean square error of cross-validation ($RMSE_{CV}$) and root mean square error of prediction ($RMSE_P$) during test validation are the criteria based on which the performance of the final models was evaluated in association with the determination coefficients (R^2_{CV} and R^2_P).

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{8}$$

$$R^2 = \left(1 - \frac{\sum [y_i - \hat{y}_i]^2}{\sum (y_i - \overline{y_i})^2}\right) \times 100 \tag{9}$$

where y_i represents the measured content of the sample, $\hat{y}_i$ is the predicted content of the sample, $\overline{y_i}$ is the mean measured content, and n represents the number of samples.

4 Experimental Validation Setup

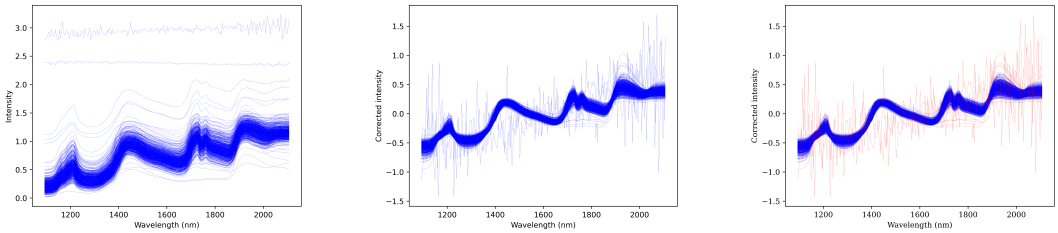
The experimental design used to validate the proposed approach targets to comprehensively explore the impact of different process parameters on the quality of the bakery products (i.e., donuts), with the ultimate goal of optimizing the process to achieve superior quality and consistency of the finished product. The extensive dataset generated through this experiment provides valuable insights into the relationships between the key process parameters and their impact on the quality attributes of donuts.

The measurements were carried out in February 2023 at the Vandemoortele Research and Development Center in Izegem, Belgium.

4.1 Samples

The full dataset is composed of 719 samples with reflectance percentage data for each of the 256 intervals between 1100-2100 nm.

The produced samples aim to assess the impact of three key factors, namely the mass concentration of shortening, proofing duration, and frying duration, on the fat content of donuts after frying. To this end, a total of 45 distinct parameter sets were designed, with each parameter set comprising varying levels of shortening mass concentration (ranging from 0% to 8% at 2% intervals), proofing duration (at 26, 38, and 50 minutes), and frying duration (at 92, 138, and 184 seconds). The frying duration is achieved by controlling the number of flips that are made by the operator during frying. As each flip takes 46 seconds, the frying duration can also be represented as 2, 3, and 4 flips.



(a) (b) (c)

Figure 1: The raw data (1a) is preprocessed by MSC-correction (1b) and outlier (red) filtering (1c)

To account for any potential variability in the production process, four donuts were produced for each of the 45 parameter sets, leading to a total of 180 donuts being evaluated in the study. The measurements were conducted at two different stages, immediately after frying and one hour after frying when the donuts were cooled. Both the top and bottom portions of each donut were measured to assess fat migration and settling during cooling. The total number of measurements obtained was 719. One sample was lost due to a technical malfunction of the software.

4.2 Sample preparation and measurement

In accordance with the predetermined experimental design, the preparation of the donut samples was carried out in a sequential manner, utilizing manual techniques that closely emulate those employed in the industrial process. The donuts were grouped into batches of twelve, with each batch ordered based on the amount of shortening fat and proofing duration.

Following frying, each batch was carefully measured on both the top and bottom sides to evaluate their respective quality attributes. Once the frying process was complete, the samples were allowed to cool down to room temperature before being subjected to a second round of measurements to ensure the consistency and reliability of the data obtained.

To prevent any interference from residual material on the sample surface, a paper towel was used to remove any remaining residue between measurements. The experimental setup and methodology used in this study were designed to ensure that the data obtained accurately reflects the true quality attributes of the donut samples, and can serve as a reliable basis for further analysis and interpretation.

4.3 NIR Hyperspectral measurement system

The samples were analyzed using the Polytec Advanced Spectrometer (PAS)-2120 NIR spectrometer equipped with the PAS-H-B01 contact sensor head. The PAS spectrometers are widely recognized as an optimal solution for NIR process analytics challenges. The devices employ diode array detector technology in combination with a transmission grating design, enabling fast and reliable data acquisition. The PAS spectrometer works in the 1100-2100 nm wavelength range with 256 subdivisions, resulting in a resolution of less than 7.9 mm. The utilization of the Polytec PAS spectrometer and its specialized contact sensor head enabled accurate and precise analysis of the samples [15].

4.4 Reference analysis

Moisture and fat content determination were performed on the same sample obtained by mixing two donuts together using a Thermomixer for 6 seconds at the maximum level. After mixing, the sample was stored in a sealable plastic bag to prevent moisture loss.

Fat content of the samples was determined using the Weibull-Stoldt method on 15 ± 0.01 g mixed sample. After acid destruction of the sample using 37 % HCl at 200 °C for 4 hours at a mixing speed of 200 rpm, the mixture was filtered using approximately 800 ml water until a neutral pH was obtained. The residue was dried overnight at 85 °C after which it was applied in a Soxhlet extraction using a 3/2 mixture of hexane and isopropanol for 6 hours at 300 °C. The extracted fat and solvent were distilled in a rotavapor until no solvent was remaining. The remaining fat was weighed and, by means of the initial sample weight, converted to a relative amount of fat in the sample.

Moisture of the sample was determined by drying 5 ± 0.001 g of mixed sample overnight at 103 ± 2 °C in a drying oven. Dried samples were cooled down for a minimum of 1 hour in a desiccator after which they were weighed again. In analogy to the fat content, the relative moisture content was calculated by dividing the dry sample weight by the initial sample weight.

5 Results

The study observed not only the desired variation in fat concentration but also variations in the crust texture, moisture content, dough distribution, and size of the donut. This variation was caused by the intentional discrepancies in the dough ingredients and processing steps, i.e. the shorting fat concentration, frying duration and proofing duration. The results initially center around an initial analysis and interpretation of the absorption spectra. Subsequently, the fat and moisture content were determined via multiple preprocessing steps and fitted to two models, namely PLS-R and LS-SVM. In the final section, the benefits of incorporating feature elimination are demonstrated through an increase in accuracy.

5.1 Interpretation of absorption spectra

The raw signals originating from 719 recorded samples during the experiment phase are plotted in Figure 1a. A large offset and magnitude difference can be seen between the signals. By implementing MSC and SNV respectively the samples could be aligned. This results in a better sample set to extract features from the data as seen in Figure 1b. The dataset contains a small number of false measurements, caused by incorrect alignment of the dough over the sensor or false triggers of the sensor. Seven outliers could be deduced by implementing thresholds to the Mean Squared Error (MSE) and Mahalanobis distance. Figure 1c shows the removed outliers compared to the dataset. The resulting outliers were identical for both MSE and Mahalanobis distance. The remaining 712 samples were used for further analysis.

Changes in the fat content of bakery dough result in alterations in the spectral absorption across multiple wavenumber ranges. Several peaks exhibit minimal variations between different sets of process parameters. For samples with zero percent shortening fat concentration, the absorption spectra grouped by frying duration, as shown in Figure 2, reveal a negative linear correlation between frying duration and intensity in the 1900-1975 nm range. The intensity shifts around 1925 nm suggesting a lower absorption for that spectral band as the frying process progresses. The entire dataset displays similar trends in different spectral bands, where a linear correlation can be observed between the process parameters and the spectral response. Again looking at Figure 2, no singular

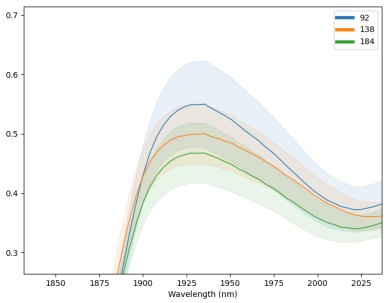


Figure 2: The average spectral response with a confidence interval of 68% for 92, 138, 184 minutes of frying with no added shortening fat. %.

wavelength clearly shows with high confidence a high linear relation. This indicates the need to extract and combine multiple indicative wavelengths to predict a single label.

5.2 Determination of fat content

The fat content of the samples was between 21.1 and 70.7 %dm, as established by the Weibull-Stoldt method. MSC and SNV were used with no additional preprocessing, next to the first derivative (FD) and second derivative (SD). In total six combinations of preprocessing techniques were implemented and compared in Table 1: MSC, SNV, MSC + FD, MSC + SD, SNV + FD, and SNV + SD.

For the regression modeling, a cross-validation strategy was adopted, by splitting the dataset into five folds and rotating the training and validation among these folds. The PLS-R model achieves a consistent accuracy over the preprocessing techniques with SNV having a small improvement over MSC. Taking the first and second derivatives for the PLS-R model again improves the accuracy by a small margin. For the LS-SVM on the other hand, the model is unable to create a valid result from the first and second derivative transformations. Overall the combination with the best result is the LS-SVM with SNV preprocessing, resulting in an R^2 and RMSE of 0.65 and 7.67 respectively for the cross-validation, although this is only a minor improvement over the PLS-R model results with SNV and second derivative.

5.3 Determination of moisture content

As a reference, the moisture content of the samples is determined by measuring the weight before and after a drying step to values between 12.1 and 33.2 %. Table 2 shows the resulting metrics for the different preprocessing techniques that are implemented. Analogous to the fat content, SNV results in a small accuracy gain over MSC for each derivative step. Similarly, the derivatives show a small improvement in accuracy over the previous step. The LS-SVM model shows no improvement on the datasets preprocessed with MSC and SVN, with again no valid fit for the derived functions. The overall best result is the SNV and second derivative fitted to the PLS regression model with a R^2 of 0.64 and RMSE of 3.11 for the cross-validation.

Table 1: Cross-validation and prediction metrics of different models and preprocessing methods for the deterimination of fat

Model	Preprocessing	Cross-validation		Prediction	
		Rc^2	RMSECV(%)	Rc^2	RMSEP(%)
PLSR	MSC	0.59	8.21	0.73	7.44
	SNV	0.60	8.09	0.73	7.48
	MSC + FD	0.61	8.03	0.69	7.93
	MSC + SD	0.62	7.82	0.69	8.05
	SNV + FD	0.61	7.99	0.69	7.94
	SNV + SD	0.64	7.70	0.70	7.78
LS-SVM	MSC	0.61	8.00	0.71	7.68
	SNV	**0.65**	**7.67**	0.70	7.80
	MSC + FD	-0.01	13.11	-0.01	14.46
	MSC + SD	-0.04	13.27	-0.04	14.63
	SNV + FD	0.14	12.07	0.17	13.08
	SNV + SD	-0.04	13.26	-0.04	14.62

Table 2: Cross-validation and prediction metrics of different models and preprocessing methods for the deterimination of moisture

Model	Preprocessing	Cross-validation		Prediction	
		Rc^2	RMSECV(%)	Rc^2	RMSEP(%)
PLSR	MSC	0.61	3.25	0.81	2.52
	SNV	0.62	3.21	0.82	2.50
	MSC + FD	0.63	3.15	0.81	2.55
	MSC + SD	0.63	3.14	0.80	2.59
	SNV + FD	0.63	3.16	0.80	2.58
	SNV + SD	**0.64**	**3.11**	0.82	2.49
LS-SVM	MSC	0.61	3.26	0.78	2.72
	SNV	0.63	3.18	0.77	2.80
	MSC + FD	-0.02	5.31	0.00	5.83
	MSC + SD	-0.05	5.37	-0.02	5.90
	SNV + FD	0.14	4.87	0.20	5.23
	SNV + SD	-0.05	5.37	-0.02	5.90

5.4 Spectral band selection

Training the regression models on the full dataset induces a risk of overfitting. The highly correlated and sparse data contains a low information density for the models to infer the predictions. The results can be further improved by employing backwards feature elimination. This is validated on the highest-scoring configurations of both models; i.e., PLS-R (SNV + SD) and LS-SVM (SNV).

As seen in Figure 3 for each of the four selected configurations the accuracy is lower when using a very low or high number of features. For instance, the fat prediction with the PLS-R (SNV + SD) after backwards feature elimination of 81 features achieves a R^2 of 0.90. This is a significant

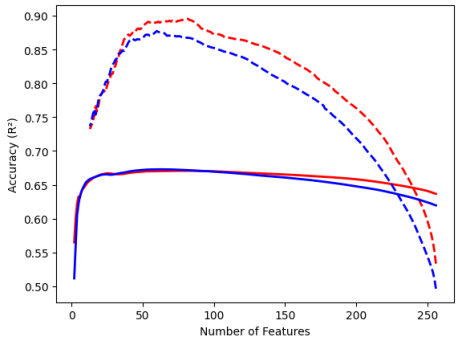


Figure 3: Accuracy of a PLS-R (dashed) and LS-SVM (full) for fat content (red) and moisture content (blue) each step in the feature elimination process

Table 3: Cross-validation and prediction metrics of PLSR (SNV + SD) and LS-SVM (SNV) for the deterimination of fat and moisture after feature selection

Label	Model (preprocessing)	Number of features	Cross-validation		Prediction	
			Rc^2	RMSECV(%)	Rc^2	RMSEP(%)
Fat	PLSR (SNV + SD)	81	0.89	4.20	0.92	3.8
	LS-SVM (SNV)	79	0.68	7.32	0.75	7.14
Moisture	PLSR (SNV + SD)	60	0.88	1.83	0.92	1.61
	LS-SVM (SNV)	63	0.68	2.96	0.83	2.39

improvement over the R^2 of 0.64 achieved with 256 features selected.

After this step, the best-achieving results of Table 1 and 2 could be improved to the values listed in Table 3.

6 Discussion

In this study, both a partial least squares regression and least squares support vector machine analysis was performed on spectral data obtained from 719 samples of donuts with varying fat and moisture contents. Several combinations of preprocessing techniques (MSC, SNV) and data transformations (first and second derivative) were evaluated in the selection of the best predictive model.

From the PLS-R results for the fat content prediction, the best determination coefficient obtained was 0.64 with SNV and second derivative, whereas LS-SVM performed slightly better than the PLS-R models with a R^2 of 0.65. Looking at the results of the determination of moisture content, the PLS-R model with SNV and second derivative results in a better determination coefficient of 0.64 in comparison to the LS-SVM with 0.63, yet again the difference is small.

Comparing the preprocessing correction techniques, SNV has an overall edge on MSC for all preprocessing techniques and models. Lastly, the derivatives do show a minor improvement with respect to the non-derived data. The first derivatives and second derivatives show an improvement of 0.02 and 0.02-0.04 of R^2 respectively.

Lastly, by incorporating backward feature elimination for the best-achieving models the accuracy increases by a significant margin. The higher accuracy increase of the PLS-R model after selecting the ideal subset of features in comparison to the LS-SVM shows that PLS-R is more prone to overfitting on high-dimensional data. After feature selection, the PLS-R gains a significant margin over the LS-SVM model.

It can be noted that the accuracy of this approach is not yet on par with the current state-of-the-art chemical method in terms of accuracy. For determining the fat content in these samples, the best approach reaches a determination coefficient of 0.89 with an RMSECV of 4.20, while the employed Weibull-Stoldt method of the reference analysis achieves an RMSECV van 1.44. Yet, the achieved accuracy can be deemed sufficient for a non-invasive primary quality control that can be executed inline.

References

[1] "Regulation (ec) no 1924/2006 of the european parliament and of the council of 20 december 2006 on nutrition and health claims made on foods," pp. L 404/9–L 404/25, 12 2006.

[2] G. Mishra, S. Srivastava, B. K. Panda, and H. N. Mishra, "Rapid assessment of quality change and insect infestation in stored wheat grain using ft-nir spectroscopy and chemometrics," *Food Analytical Methods*, vol. 11, pp. 1189–1198, 4 2018.

[3] T. Szigedi, M. Dernovics, and M. Fodor, "Determination of protein, lipid and sugar contents of bakery products by using fourier-transform near infrared spectroscopy," *Acta Alimentaria*, vol. 40, pp. 222–229, 2011.

[4] L. Gomes, T. Nobre, A. Sousa, F. Rei, and N. Guiomar, "Hyperspectral reflectance as a basis to discriminate olive varieties—a tool for sustainable crop management," *Sustainability 2020, Vol. 12, Page 3059*, vol. 12, p. 3059, 4 2020.

[5] H. Jiang, X. Jiang, Y. Ru, J. Wang, L. Xu, and H. Zhou, "Application of hyperspectral imaging for detecting and visualizing leaf lard adulteration in minced pork," *Infrared Physics & Technology*, vol. 110, p. 103467, 11 2020.

[6] N. Gat, "Imaging spectroscopy using tunable filters: a review," *Wavelet Applications VII*, vol. 4056, pp. 50–64, 4 2000.

[7] G. Polder, G. W. V. D. Heijden, L. C. Keizer, and I. T. Young, "Calibration and characterisation of imaging spectrographs," *http://dx.doi.org/10.1255/jnirs.366*, vol. 11, pp. 193–210, 6 2003.

[8] J. Burger and P. Geladi, "Hyperspectral nir image regression part i: Calibration and correction," *Journal of chemometrics*, vol. 19, pp. 355–363, 2005.

[9] H. Martens, S. A. Jensen, and P. Geladi, "Multivariate linearity transformation for near-infrared reflectance spectrometry," 1983, pp. 205–234.

[10] R. J. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra," *http://dx.doi.org/10.1366/0003702894202201*, vol. 43, pp. 772–777, 7 1989.

[11] P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," *Analytica Chimica Acta*, vol. 185, pp. 1–17, 1 1986.

[12] A. Lorber, L. E. Wangen, and B. R. Kowalski, "A theoretical foundation for the pls algorithm," *Journal of Chemometrics*, vol. 1, pp. 19–31, 1 1987.

[13] A. Croguennoc, J. Lallemand, and R. Sylvie, "Some aspects of svm regression : an example for spectroscopic quantitative predictions," 2018.

[14] J. A. Suykens, L. Lukas, and J. Vandewalle, "Sparse approximation using least squares support vector machines," *Proceedings - IEEE International Symposium on Circuits and Systems*, vol. 2, 2000.

[15] "Pas-h-b01 and pas-h-b04 contact sensor heads - polytec." [Online]. Available: https://www.polytec.com/eu/process- analytics/products/probes-for-process-spectrometers/pas-h-b01-and- pas-h-b04-contact-sensor-heads